AFRL-VS-HA-TR-2005-1113

Development of Ionospheric Assimilation and Forecasting System

Boris Khattatov
Michael Murphy
Marianna Gnedin
Tim Fuller-Rowell
Valery Yudin
Jason Boisvert
Jeff Sheffell

Fusion Numerics, Inc.
Environmental Research Technologies
1320 Pearl Street
Boulder, CO 80302-5287

22 July 2005

Scientific Report No. 1

AIR FORCE RESEARCH LABORATORY
Space Vehicles Directorate
29 Randolph Road
AIR FORCE MATERIEL COMMAND
Hanscom AFB, MA 01731-3010

This technical report has been reviewed and is approved for publication.

AFRL-VS-HA-TR-2005-1113

/signed/

KIMBERLEE GRUENSTEIN, 1Lt.
Contract Manager

/signed/

JOEL B. MOZER, Chief
Space Weather Center of Excellence

This report has been reviewed by the ESC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center (DTIC). All others should apply to the National Technical Information Service.

If your address has changed, if you wish to be removed from the mailing list, or if the addressee is no longer employed by your organization, please notify AFRL/VSIM, 29 Randolph Rd., Hanscom AFB, MA 01731-3010. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
22-07-2005	Scientific Report No. 1	May 2004 - May 2005

4. TITLE AND SUBTITLE

Development of Ionospheric Assimilation and Forecasting System

5a. CONTRACT NUMBER
FA8718-04-C-0010

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
65502F

6. AUTHOR(S)
Dr. Boris Khattatov

Dr. Michael Murphy, Dr. Marianna Gnedin, Dr. Tim Fuller-Rowell, Jason Boisvert, Jeff Sheffell

5d. PROJECT NUMBER
3005

5e. TASK NUMBER
SD

5f. WORK UNIT NUMBER
4K

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Fusion Numerics, Inc.
Environmental Research Technologies
1320 Pearl Street
Boulder, CO 80302-5287

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
29 Randolph Road
Hanscom AFB, MA 01731-3010

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/VSBXP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-VS-HA-2005-1113

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report was developed under Small Business Innovation Research (SBIR) contract for Topic AF03-016. This interim report describes current progress in developing an ionospheric modeling and forecasting system sponsored by AFRL. We outline design of the physics-based numerical ionospheric model that serves as a core of the system and the data assimilation framework for constraining the model with GPS measurements of ionospheric electron content. A newly developed scheme for estimating GPS receiver inter-channel code biases on-line is presented. The report also describes SOPA/XML web-services infrastructure for delivering system simulation results to remote clients. Recently emerged protocol for delivering differential GPS data over the Internet, NTRIP, resulted in availability of free sources of real-time GPS streaming data. We describe development efforts for obtaining and utilizing such GPS data. We conclude by outlining a scintillation nowcasting module and challenges related to longer-term forecasting.

15. SUBJECT TERMS
Keywords: **Ionosphere, GPS, NTRIP, Data assimilation, TEC, Electron content, Forecasting, modeling, SOAP/XML, Web services**

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Lt. K. Gruenstein
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR	40	19b. TELEPHONE NUMBER (Include area code) 781 377-3107

Contents

Figures

Acknowledgments

We are grateful to the Air Force Research Laboratory personnel, in particular, to John Retterer, Odile de la Beaujardiere, Dwight Decker, William Borer, Jason Mellein, Craig Baker, and Kimberlee D. Gruenstein for guidance and continuing support of this effort. The Air Force Research Laboratory funded this project.

F.T. Berkey (USU's Bear Lake Observatory), J.W. Wright and N.A. Zabotin (both at CIRES/NGDC) provided the dynasonde data. M A. Angling (QinetiQ Ltd), J. L. Chou and D. Hysel helped to validate the model with independent data. The Jicamarca Radio Observatory is operated by the Instituto Geofisico del Peru, Ministry of Education, with support from the National Science Foundation through cooperative agreement ATM-0432565 with Cornell University.

1. INTRODUCTION

This report describes development of an operational system for generating, forecasting, and distributing three-dimensional ionospheric electron densities and corresponding Global Positioning System (GPS) propagation delays. The system adapts data assimilation technologies developed and routinely used for operational tropospheric weather forecasting in order to nowcast and forecast ionospheric conditions.

The core ionospheric model solves plasma dynamics and composition equations governing evolution of density, velocity, and temperature for 7 ion species on a fixed global three-dimensional grid. It uses a realistic model of the Earth's magnetic field and solar indices obtained in real time from the NOAA Space Environment Center. While the core model is capable of delivering realistic results, its accuracy can be significantly improved by employing a special set of numerical techniques known as data assimilation. In the process of data assimilation, the core ionospheric model is continuously fed observational data from a network of reference GPS ground stations. This improves both the nowcast and the forecast. Public access to the system is provided at http://www.fusionnumerics.com/ionosphere.

System results are validated via systematic comparisons with GPS station data withheld from the assimilation process. It is shown that globally average RMS error is approximately 5 total electron content (TEC) units and RMS error for locations in the Northern Hemisphere is approximately 3 TEC units.

Since the summer of 2004, model output after assimilation has been supplied to the National Geophysical Data Center and compared with data from the Bear Lake (Utah) dynasonde. These systematic comparisons have been crucial to the development of the system and used extensively for its validation. We present sample statistics of the system-derived plasma frequencies and the dynasonde observations. We also present comparisons with electron density profiles measured at the Jicamarca observatory.

This report also describes the procedure for obtaining electron densities from our operational assimilation system remotely via the internet.

2. METHODS, ASSUMPTIONS, AND PROCEDURES

2.1. Ionospheric Model

The developed model (Khattatov et al. 2004, a, b, c) is a numerical global model of the ionosphere loosely based on descriptions given in Bailey and Balan (1996), Fuller-Rowell et al. (1996), Millward et al. (1996) and Huba at al. (2000). It consists of over 90,000 lines of C++ code, is fully 64-bit compliant, adheres to standard C++ language specifications, and was ported to 64-bit and 32-bit Linux and UNIX platforms and MS Windows. The code has been developed using modern software engineering tools and methodologies in an object-oriented fashion. The model computes the spatial distribution and temporal evolution of H+, O+, He+, O_2+, NO+, N_2+, N+ and electrons. The model solves momentum and mass conservation equations for all seven ion species and electrons, and the energy conservation equation for the three major ions and electrons. It includes chemical interactions with neutrals and ions, recombination, ion-ion and ion-neutral collision rates, photoionization, and different types of heating.

The model domain covers all latitudes and longitudes; however, implementation of polar transport and high-latitude effects are still in an experimental stage and are not considered reliable. As customary in ionospheric applications, the dynamic equations are solved in so-called magnetic coordinates since, in the absence of electric fields, plasma moves predominantly parallel to the direction of the magnetic field. A detailed discussion of the coordinate transformation and related equations can be found, for instance, in Millward et al. (1996), and Bailey and Balan et al. (1996). Here, we give only a brief overview for the benefit of readers not familiar with this subject.

The Earth's magnetic field is approximated as that of a tilted eccentric dipole. Model equations are given in dipole coordinates, along magnetic field lines. The first magnetic coordinate is magnetic longitude. For each magnetic longitude, we consider a "vertical stack" of magnetic field lines characterized by the distance of the apex of each line from the Earth's center at the magnetic equator. This distance, normalized by the Earth's radius, is the second magnetic coordinate, p. Finally, for each field line, the distance from the apex to a particular point along the magnetic field line gives the third coordinate, s. A portion of the model grid is shown in Figure 1.

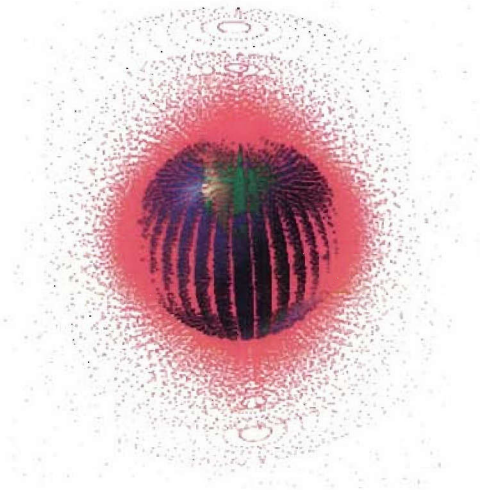


Figure 1. A portion of the model spatial grid. Only every 8th point is shown for clarity.

Once the field-aligned transport is solved, we compute plasma evolution due to cross-field transport. Cross-field transport is forced by electric fields either imposed externally from the magnetosphere or generated internally from the action of the neutral wind. In the lower thermosphere, the mobility of the ions is inhibited by collisions with the neutral atmosphere. The dynamo action of the neutral winds drives currents that create polarization electric fields through continuity. The ions respond to these electric fields by drifting perpendicular to both the electric and magnetic fields. This is often referred to as **ExB** transport. In non-fully coupled models, the **ExB** plasma velocity is specified from external empirical models. Once this velocity is known, solving the plasma advection equation is relatively straightforward. To model **ExB** drift at low latitudes, we use coefficients of the Fejer and Schierless model from International Reference Ionosphere 2001. To model **ExB** drift in the polar regions, we use coefficients of the Weimer (1996) model provided by Dr. Daniel Weimer of Mission Research, Inc.

2.2. Data Assimilation

The Kalman filter is a mathematical tool traditionally used to continuously combine model predictions of the system state with noisy observations to produce a less uncertain estimation of the system state with error covariances. The adopted Kalman filter approach is derived from Khattatov et al. (2000). Let us assume that model estimates of electron densities at all grid points at time t are arranged in a vector $\mathbf{x}$ with dimension N_x. Formally, integration of the model $\mathbf{M}$ can be written as

$$\mathbf{x}_{t+\Delta t} = \mathbf{M}(t, \mathbf{x}_t) \tag{1}$$

Let vector $\mathbf{y}$ contain observations of a quantity linearly related to electron densities at the same time. In the case of GPS reference-station data, such quantities are slant total electron content (TEC) from each station to all satellites in view. The connection between $\mathbf{x}$ and $\mathbf{y}$ is established via linear observational operator $\mathbf{H}$ as follows:

$$\mathbf{y} = \mathbf{H}(\mathbf{x}) \tag{2}$$

Under assumptions of linearity and Gaussian statistics, the optimal value of $\mathbf{x}$ that inverts Equation (2) given a set of observations $\mathbf{y}$ and model estimates of $\mathbf{x}$ is given by:

$$\mathbf{x}_t^a = \mathbf{x}_t + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_t) \tag{3}$$

$$\mathbf{K} = \mathbf{B}_t\mathbf{H}^T(\mathbf{H}\mathbf{B}_t\mathbf{H}^T + \mathbf{O} + \mathbf{R})^{-1} \tag{4}$$

Here, $\mathbf{B}_t$ is the forecast error covariance at time t, $\mathbf{O}$ is the error covariance matrix of the observations, and $\mathbf{R}$ is the representativeness error covariance associated with errors of interpolation and discretization. Matrix $\mathbf{K}$ is called the Kalman gain matrix.

The analysis error covariance is expressed as:

$$\mathbf{B}_t^a = \mathbf{B}_t - \mathbf{B}_t\mathbf{H}^T(\mathbf{H}\mathbf{B}_t\mathbf{H}^T + \mathbf{O} + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}_t \tag{5}$$

Once inversion of Equation (2) is performed, the obtained electron densities, $\mathbf{x}_t^a$, can be used as the initial condition for the model $\mathbf{M}$ to predict electron densities at a later time (beginning of the next assimilation window) according to Equation (1):

$$\mathbf{x}_{t+\Delta t} = \mathbf{M}(t, \mathbf{x}_t^a) \tag{6}$$

Since the model domain contains about 10^6 points, direct matrix manipulations described by Equations (3), (4), and (5) are impossible to implement even with modern computing capabilities. As in Khattatov et al. (2000), only the diagonals of the updated error covariance matrix are computed from Equation (1). The off-diagonals are parameterized as 3-D separable Gaussian models with prescribed de-correlation lengths. The values of these parameters are later tuned to minimize post-fit residuals and satisfy the χ^2 test described in Khattatov et al. (2000).

The time-dependent differential equations governing evolution of the ionospheric plasma are given in the magnetic coordinate system aligned with the constant magnetic field lines of the Earth. This is also the internal coordinate system of the numerical ionospheric forward model. It is therefore natural to

specify background error covariances or, in our implementation, de-correlation lengths, in this magnetic coordinate system.

Moreover, it is highly desirable to implement the analysis Equations (3) and (4) in the magnetic coordinate system rather than the commonly used geographic coordinate system, even though the measurement geometry is more naturally described in the geographic system. This avoids double interpolation of the model state from magnetic to geographic system prior to the analysis step and back after the analysis. Such interpolation will have introduced additional errors at each assimilation interval. Ionospheric conditions often change rapidly at time scales of minutes and GPS measurements are typically available as often as every second. It is therefore not unreasonable to expect an assimilation analysis performed every several minutes. The double interpolation procedure has a potential for introducing a substantial error at each analysis step, therefore, possibly making frequent updates useless or even detrimental to the system's performance.

Following these arguments we implemented the analysis equations in the magnetic coordinate system. To compute observational operator **H** in magnetic coordinates we start by "drawing" a straight line between a given receiver and a satellite, and breaking it into a large number of discrete points. For each point we identify the two closest p-q slices and compute Delaunay triangulation for each 2-D p-q slice. We then project the position of the point to each p-q plane and compute interpolation weights for the 6 surrounding model grid points from the pre-computed Delaunay triangulations. These weights and indices of the interpolating model gridpoints are stored for each point on the receiver-satellite line of sight. The procedure is then repeated for all points and the line-of-sight integral is computed. Since the same model gridpoints can contribute to interpolation for multiple segments on the line of sight, a special algorithm searches for such instances and consolidates interpolation weights so that any model gridpoint is only included once in the observational operator. These consolidated weights become elements of matrix **H** in Equation (2).

2.2.1. Slant TEC Determination from the Global Positioning System Data

The space segment of the Global Positioning System (GPS) (Parkinson et al. 1996) is a constellation of (at the time of this writing) 29 satellites located on orbits about 20,000 km above the Earth's surface. At any given time, at least 4 satellites are visible from most locations on Earth. Each GPS satellite emits a signal that contains a unique identification code (pseudo random number, PRN), accurate time from an on-board atomic clock, satellite position, and auxiliary information. A GPS receiver that has 4 or more GPS satellites in view can, by "triangulation" in space and time, determine both precise time and position.

GPS L-band frequency signals are delayed by the ionosphere approximately proportional to the total integrated electron content along the line of sight between a receiver and a GPS satellite. These delays can result in positional errors of tens of meters. To mitigate this effect, GPS satellites and high-end GPS receivers use signals on 2 different frequencies referred to as L1 and L2. Since the ionosphere is a dispersive media at these frequencies, it is possible to estimate and remove the ionospheric delays and therefore the slant total electron content (TEC) for each dual-frequency receiver-satellite pair.

The ionospheric delays can be estimated by subtracting the perceived distances between the receiver and the satellite (pseudoranges) measured at the two frequencies. In the absence of other systematic errors this would result in the absolute albeit noisy estimate of the total electron delay (see Figure 2). A GPS receiver can also very accurately measure a phase shift between received satellite carriers at the two frequencies and receiver-generated replica of the satellite carrier. These measurements can be converted to a relative change of slant TEC between a given receiver and the satellite as a function of time (Figure 2). If the receiver loses track of the satellite due to decreased signal-to-noise ratio or other

interferences, the phase-generated TEC will experience a sharp jump called cycle slip. A relatively standard approach in GPS data processing (e.g., Blewitt 1990) is to identify and correct for cycle slips and then add a smoothed shift between phase- and pseudorange-derived TEC to the phase measurements.

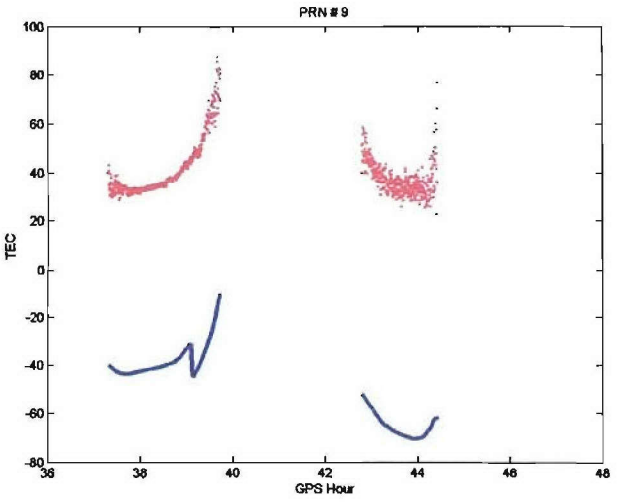


Figure 2. An example of relative slant total electron content (TEC) between IGS station ARTU and GPS satellite designated PRN#9 as a function of time. Dots represent pseudorange-derived TEC which is generally noisy. Solid line corresponds to phase-derived TEC.

International GNSS Service (IGS) operates a network of dual frequency GPS receivers with data available to the public at no charge albeit often with significant time delays (up to 2 days). GPS stations used in our system for determining the slant electron content are shown in Figure 3.

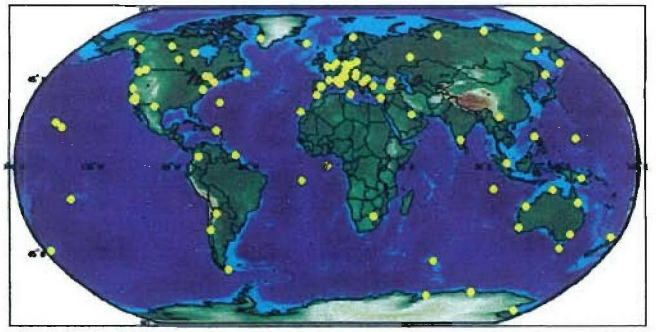


Figure 3. Locations of IGS stations used in the system.

Unfortunately slant TEC estimates obtained in this fashion are still not free of systematic errors. These satellite and receiver specific errors arise from the fact that L1 and L2 signals take slightly different times to travel from the radio to the antenna (on a satellite) and from the antenna to the receiver. Receiver's circuitry can introduce additional differential signal delays. Thus, each estimated slant TEC value between a receiver and a satellite contains a systematic error equal to the sum of the receiver and the satellite hardware biases. These biases are often much larger than the "true" TEC value. The satellite biases are generally better known and are more stable. Receiver biases vary, albeit slowly, due to changes in environmental conditions, equipment aging, and possibly other factors.

Clearly, for the reference station data to be useful such biases need to be estimated, monitored, and removed from the measurements.

Traditionally, these biases are estimated in the off-line mode after accumulating a time series of measurements from a number of reference stations (e.g., Mannucci et al. 1998). In the next section, we introduce an approach to on-line receiver hardware bias estimation using the same Kalman filter employed in the state estimation but augmented with the bias unknowns.

2.3. Bias Determination

Dual-frequency GPS receivers normally have large differential code biases (DCB). These biases change with time, often at a rate of 1-2 TEC units per day. Much larger abrupt changes are possible due to equipment (e.g., antennae, cables) maintenance or replacement. All this makes it highly desirable to be able to estimate these biases in near real time rather than in the post processing mode. We implemented the equipment bias estimation scheme by explicitly introducing a vector of hardware biases **b**:

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{b} \tag{7}$$

In Equation (7), the "hat" modifier denotes unbiased observations. We augment the state vector **x** with the set of biases and solve the new analysis Equation (8) for the additional unknowns:

$$\hat{\mathbf{x}} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \\ \dots \\ b_1 \\ b_2 \\ b_3 \\ \dots \end{bmatrix} \qquad \hat{\mathbf{x}}_t^a = \hat{\mathbf{x}}_t + \hat{\mathbf{K}}(\mathbf{y} - \hat{\mathbf{H}}\hat{\mathbf{x}}_t) \tag{8}$$

The new observational operator $\hat{\mathbf{H}}$ now contains extra columns containing zeroes almost everywhere and 1s at the raw corresponding to the given receiver bias.

In the absence of spatial correlation in the background error covariances, Equation (8) will estimate the biases by taking an average difference between the model and the data. The model itself however contains its own biases due to misrepresentation of physical processes and numerical errors. In general, model biases will be a function of time and space. Therefore, unless the model biases are negligible, Equation (8) cannot yield a reasonable estimate of observational biases.

It is the presence of spatial correlation in the background model field that makes bias estimation via Equation (8) possible. Lines of sight between two nearby receivers and some GPS satellites often pass through the same region of the ionosphere. In the extreme case of 2 co-located receivers measuring slant TEC to the same satellite(s), one can readily compute the difference between their respective hardware biases thus adding an extra constraint to the system. Non-zero de-correlation lengths in the

Kalman filter equations effectively allow similar constraints for non co-located receivers. It is possible that over time, for a network of receivers constantly measuring slant TEC to all orbiting GPS satellites, there will be enough information to estimate both the hardware biases and the ionospheric state. Other constraints facilitating bias estimation are due to the fact that at nighttime ionospheric total electron content is very close to zero due to the absence of solar radiation and the absolute nighttime model biases are rather small.

Results of equipment bias estimation with the described procedure are illustrated in Figure 3 for several GPS receivers. After an initial adjustment period that can last a couple of days (not show in the figure), the estimated biases become relatively stable. It is difficult to conclude whether the remaining time variability is due to non-optimal filtering or the actual bias changes in the receiver. For comparison, we also included independent bias estimates by Jet Propulsion Laboratory International GPS Analysis Centre. Both estimates typically agree to within 1-5 TEC units which is close to the usual bias changes reported by the JPL IGS Centre.

Note that JPL-estimated bias values change daily and thus can experience large jumps at the day boundaries (see Figure 4).

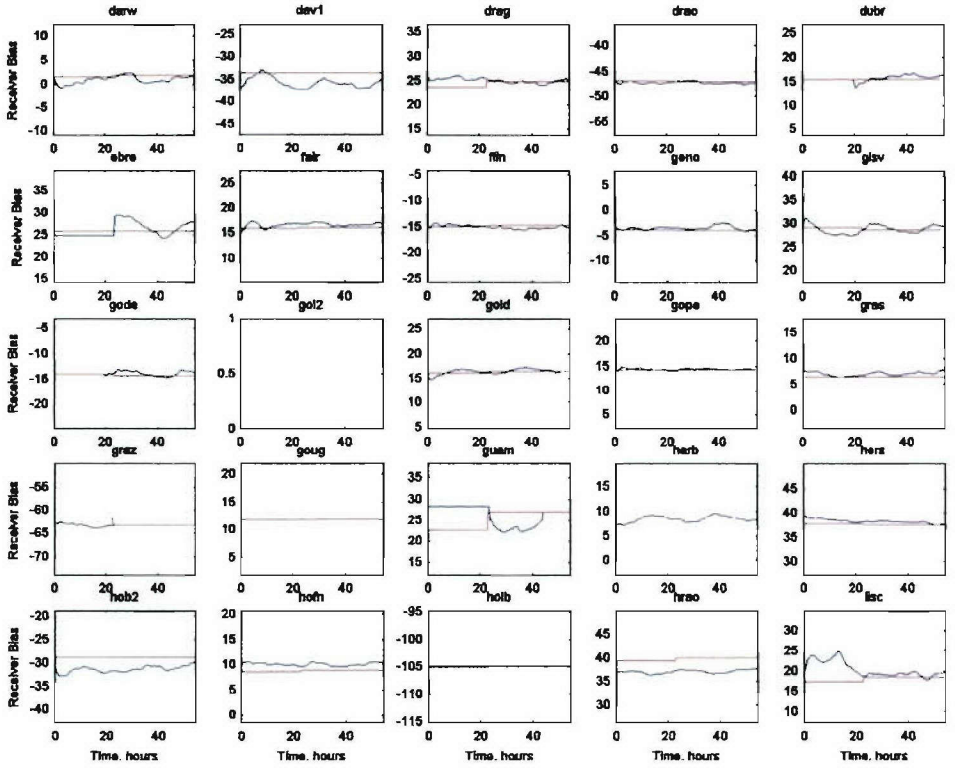


Figure 4. An example hardware bias estimation for several different GPS receivers in TEC units as a function of time in hours. The flat lines correspond to post-processed biases estimated by the Jet Propulsion Laboratory. The oscillating lines are the results of the dual state-bias filter estimation procedure.

2.4. Data Delivery Mechanism

The ionospheric specification model is currently run operationally in 2 configurations. In the first configuration, the system is synchronized with real time but does not perform any data assimilation due to latencies associated with IGS network data. In the second configuration, the system is synchronized

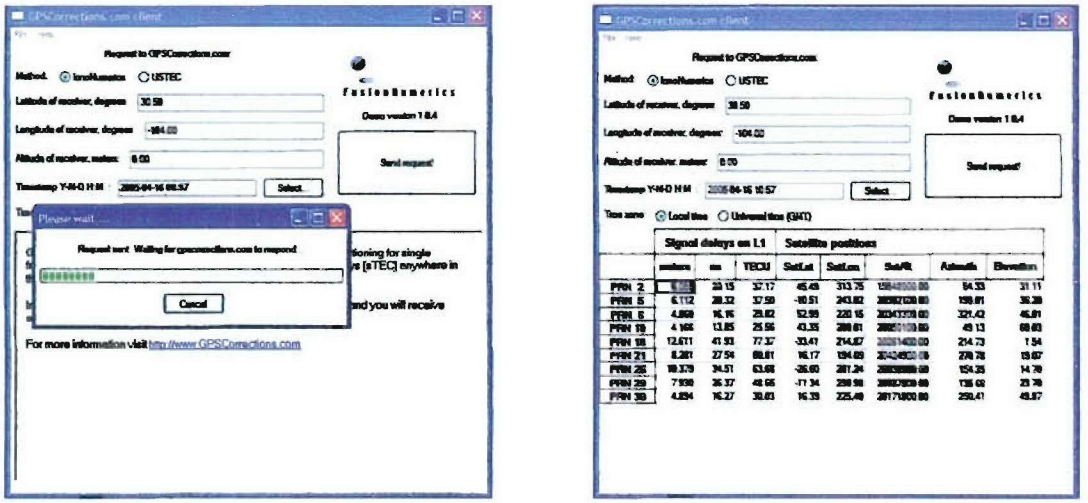
with real time minus 72 hours and includes the data assimilation module. Ionospheric parameters generated in the system such as slant or vertical total electron content, local electron densities and temperatures, etc., can be obtained by a casual end user in the form of color coded plots superimposed on the map of the Earth from http://FusionNumerics.com/ionosphere . Additionally, the user can download hourly 3-D data files in platform independent netCDF format from the web page.

For more demanding or automated applications, we have implemented a remote procedure call mechanism based on the web services standard. A web service is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-readable format. Other systems (clients) interact with the web service in a manner prescribed by its description using SOAP (Simple Object Access Protocol) messages, typically conveyed using HTTP with an XML (eXtendable Markup Language) serialization. Web services is a modern, industry standard method for providing remote access to enterprise applications and data. A variety of free libraries exist for practically all common programming languages that allow a remote client to consume any exposed web service.

The client and the server exchange information in XML format over the internet http protocol using POST command. The end user does not have to know the specifics of the XML format or SOAP as freely available libraries require writing of only a few lines of code to consume the web service. The developed web service, residing on the backend Linux server, listens for incoming SOAP/HTTP requests, verifies their validity, and automatically returns user-requested data (slant TEC values from the user's location to all GPS satellites in view). The code fragment below shows a portion of the working web-services client written in popular scripting language Perl:

```
# Create the SOAP client

my $client = SOAP::Lite->new();

# Specify the name of the web service

$client->uri('urn: getTEC');

# Specify the url of the server

$client-proxy('http://FusionNumerics.com/ionosphere/webservices');

# Call the SOAP server; Perl array MyTecRequest  contains time and latitude an longitude of
# the location for which vertical TEC is  requested

my $response = $client->Tec(@MyTecRequest);

# Assign the SOAP response; Perl variable
# TecResult now contains vertical TEC

my @TecResult = @{$response->result}
```

Similar clients can be easily created in C, Java, etc. Further specifics as well as demo lightweight clients in Perl, C and C++ can be obtained from the authors. These clients can then be used to remotely access our "live" server and deliver estimated slant TECs for a variety of locations and times. Perl and C source code web clients, for this web service, are available from Fusion Numerics.



```
linux> more input.txt
40 -105 100 000000000000
30 -105 100 000000000000
20 -105 100 000000000000
10 -105 100 000000000000

linux> ./TecSoapClient < input.txt > output.txt

linux> more output.txt
Sending 4 requests to server...
SOAP Server response received
Response Count: 4
------ Response to request 1
TECs for 10 visible satellites (Lat Lon Alt Az El)
  TEC for PRN  3:  44.77 ( 40.47  190.40 20200000.00  292.80   29.77)
  TEC for PRN  9:  42.38 ( 17.28  305.50 19860000.00  102.30   28.75)
  TEC for PRN 14:  50.43 ( -4.20  232.40 20190000.00  210.70   29.27)
  TEC for PRN 15:  25.89 ( 48.56  247.20 19970000.00  329.90   76.41)
  TEC for PRN 18:  25.97 ( 55.03  262.70 20250000.00   16.20   69.07)
  TEC for PRN 19:  55.77 ( 54.35  160.80 20280000.00  318.20   15.84)
  TEC for PRN 21:  29.66 ( 24.09  278.70 19960000.00  121.20   57.02)
  TEC for PRN 22:  29.50 ( 44.95  222.40 20300000.00  292.80   58.37)
  TEC for PRN 26:  46.09 ( 54.77  346.80 19980000.00   42.10   17.24)
  TEC for PRN 29:  52.98 ( 48.67    4.95 20430000.00   40.76    4.18)
------ Response to request 2
TECs for 11 visible satellites (Lat Lon Alt Az El)
  TEC for PRN  1: 131.60 (-32.34  211.90 20100000.00  216.80    1.73)
  TEC for PRN  3:  56.92 ( 40.47  190.40 20200000.00  300.20   24.73)
  TEC for PRN  6: 116.70 (-36.04  265.80 20040000.00  170.50    9.63)
  TEC for PRN  9:  49.18 ( 17.28  305.50 19860000.00   93.49   30.32)
  TEC for PRN 14:  48.81 ( -4.20  232.40 20190000.00  216.30   38.93)
  TEC for PRN 15:  31.97 ( 48.56  247.20 19970000.00  344.50   64.39)
  TEC for PRN 18:  32.96 ( 55.03  262.70 20250000.00   10.21   56.69)
  TEC for PRN 19:  78.54 ( 54.35  160.80 20280000.00  321.30    7.81)
  TEC for PRN 21:  33.66 ( 24.09  278.70 19960000.00   99.82   61.50)
  TEC for PRN 22:  36.25 ( 44.95  222.40 20300000.00  309.60   51.84)
  TEC for PRN 26:  64.22 ( 54.77  346.80 19980000.00   38.83    9.17)
------ Response to request 3
TECs for 11 visible satellites (Lat Lon Alt Az El)
  TEC for PRN  1: 129.50 (-32.34  211.90 20100000.00  219.00    9.78)
  TEC for PRN  3:  73.78 ( 40.47  190.40 20200000.00  306.00   18.73)
  TEC for PRN  6: 107.80 (-36.04  265.80 20040000.00  169.60   20.09)
  TEC for PRN  9:  63.41 ( 17.28  305.50 19860000.00   84.39   30.10)
  TEC for PRN 14:  53.51 ( -4.20  232.40 20190000.00  225.10   48.03)
  TEC for PRN 15:  43.16 ( 48.56  247.20 19970000.00  349.40   52.14)
  TEC for PRN 18:  45.89 ( 55.03  262.70 20250000.00    7.61   44.58)
  TEC for PRN 19:  92.52 ( 54.35  160.80 20280000.00  323.20   -0.21)
  TEC for PRN 21:  42.49 ( 24.09  278.70 19960000.00   74.89   60.87)
  TEC for PRN 22:  48.49 ( 44.95  222.40 20300000.00  320.40   43.22)
  TEC for PRN 26:  77.65 ( 54.77  346.80 19980000.00   36.72    1.11)
------ Response to request 4
TECs for 13 visible satellites (Lat Lon Alt Az El)
  TEC for PRN  1:  97.77 (-32.34  211.90 20100000.00  222.40   17.80)
  TEC for PRN  3:  98.79 ( 40.47  190.40 20200000.00  310.30   12.14)
  TEC for PRN  6:  80.52 (-36.04  265.80 20040000.00  168.00   31.01)
  TEC for PRN  9:  85.13 ( 17.28  305.50 19860000.00   75.73   28.13)
```

Figure 5. Screenshots of Microsoft Windows (top) and Linux client invocation and responses. Linux input file, input.txt, contains four requests for latitudes of 40, 30, 20, and 10 degrees, longitude of 105 degrees and altitude of 100 m. Time stamp of 000000000000, optionally specified as DDMMYYYYHHMM, corresponds to a real time request.

An example of the client's request and the server's response are shown in Figure 5 for the Microsoft Windows GUI client and a command line Linux client. The client simply makes one or more requests

for a particular latitude, longitude, and time, and receives one or more server responses that shows slant TEC values and L1 signal delays to visible satellites.

Depending upon which of the various web service *methods* are called, the associated method calculation results are returned to the user via the web service.

The currently implemented Web Service methods are:

Tec - Slant TECs to visible satellites (default method),

Lec - Local Electron Content,

Ptpec - Point-to-point Electron Content, and

Ustec - Slant TECs to visible satellites for CONUS from NOAA SEC USTEC model.

For example, if a query is made to the *Tec* method (Total Electron Content), then *Slant TEC* values, to every visible GPS satellite, from each requested earth-relative position, are returned. The integrated electron densities, from satellites to receiver positions, are known as the Slant Total Electron Content or Slant TEC.

Refer to the following websites for more details: http://www.iononumerics.com/ .The latest Web Service Description Language (WSDL) document is located here:
http://69.15.204.66:8080/iononumerics/IonModel-Tec.wsdl

2.4.1. Using the Web Service via a SOAP Client

Refer to individual documentation (usually a README file) that accompanies each SOAP client for details on usage of that particular client. Such details include command line options for the particular client (e.g., cSoap or pSoap clients).

An archive of recent model output data is available for access via web service queries. There is a time period, into the past, for which the archive data is made available. The time period moves along with current time, creating a sliding historical time window for which the system output is available.

Additionally, there is a near real time model that does not perform data assimilation. The real time model is a physics-based model, with no GPS data input.

The web service does not distinguish between the real time and the assimilation models. The user will get results based upon the current state of the model output files, with respect to the user's requested timestamps. The *current timestamp feature*, in which an all-zero timestamp is specified in the user's request (i.e., "000000000000"), will return results from the most current IonoNumerics model output data.

2.4.1.1 Tec Method – Slant Total Electron Content to Each GPS Satellite

A single SOAP request contains multiple receiver positions, each with an associated timestamp. The web service returns, for each position requested, the Slant TEC calculations for each visible satellite (or visible PRN). The receiver positions are specified as earth-relative latitude, longitude, and altitude.

Input Format:

ReceiverLat receiver latitude, +-90 degrees, floating point

ReceiverLon receiver longitude, +-180 degrees, floating point

ReceiverAlt receiver altitude in meters, from sea level, -500 to 800,000 meters, floating point

Time requested time of calculation, DDMMYYYYHHMM or 000000000000,

Each single line contains the input parameters. Latitudes and longitudes are double precision values with N digits of precision.

Output Format:

ReceiverLat receiver latitude

ReceiverLon receiver longitude

ReceiverAlt receiver altitude

Time requested time of calculation

PRNCount number of visible satellites, integer

ErrorCode return error (refer to error table), integer

For each input line, a record is returned. In summary, the header line repeats the input parameters and adds the PRNCount and ErrorCode. Only the slant TECs for each visible satellite are returned. For each header, a line is output for each visible satellite with the following format:

PRN PRN number of the visible satellite, integer

SlantTEC calculated Slant TEC for the satellite, floating point

SatLat latitude of the satellite, floating point

SatLon longitude of the satellite, floating point

SatAlt altitude of the satellite in meters, floating point

SatAz satellite azimuth angle from receiver

SatEl satellite elevation angle from receiver

For multiple receiver position requests, within a single SOAP request, the response will be ordered the same as the request order.

If the near real time option is specified, then the timestamps will be returned as an all-zero field ("000000000000").

2.4.1.2. Ustec Method - U.S. Total Electron Content

The *Ustec* web services method accesses data from NOAA Space Environment Center (SEC) USTEC ionospheric model. Yet, the user or web services interface and functionality are the same as the Tec method.

The Ustec method confines the availability of model output data to earth positions within the continental United States (Conus). User requests outside of the Conus *bounding box* will return zero visible satellites.

A single SOAP request contains multiple receiver positions, each with an associated timestamp. The web service returns, for each position requested, the Slant TEC calculations for each visible satellite (or visible PRN). The receiver positions are specified as earth-relative latitude, longitude, and altitude.

The input specifications for the Ustec method are the same as the inputs for the Tec method, except that receiver altitude is not used for the Ustec method, although it is a required input parameter. Since the altitude is ignored, any value may be input.

Input Format:

ReceiverLat receiver latitude, +-90 degrees, floating point

ReceiverLon receiver longitude, +-180 degrees, floating point

ReceiverAlt while altitude is not used, it is required, floating point

Time requested time of calculation, DDMMYYYYHHMM or 000000000000,

Each single line contains the input parameters. Latitudes and longitudes are double precision values with N digits of precision. See example input file below.

The outputs of the Ustec method are the same as the outputs for the Tec method. Refer to the Tec method Output Specifications.

2.4.1.3 Lec Method - Local Electron Content

A single SOAP request contains multiple positions, each with an associated timestamp. The web service returns, for each position requested, the Local Electron Content (LEC) at that position. The positions are specified as earth-relative latitude, longitude, and altitude.

Input Format:

Lat point latitude, +-90 degrees, floating point

Lon point longitude, +-180 degrees, floating point

Alt point altitude in meters, from sea level, 100000 to 10,000,000 meters, floating point

Time requested time of calculation, DDMMYYYYHHMM or 000000000000,

Each single line contains the input parameters. Latitudes and longitudes are double precision values with N digits of precision.

Output Format:

Lat point latitude

Lon point longitude

Alt point altitude

Time requested time of calculation

ErrorCode return error (refer to error table), integer

For each input line, a record is returned. In summary, the header line repeats the input parameters and adds the ErrorCode.

Latitudes and longitudes are double precision values with N digits of precision. For each header, a single line is output that contains the LEC, with the following format:

Lec calculated Local Electron Content, floating point

For multiple position requests, within a single SOAP request, the response will be ordered the same as the request order.

2.4.1.4. Ptpec Method - Point-to-point Electron Content

A single SOAP request contains multiple position pairs, each with an associated timestamp. The web service returns, for each position pair requested, the Point-to-point Electron Content (Ptpec) between the two points. The positions pairs are specified as earth-relative latitudes, longitudes, and altitudes.

Input Format:

P1Lat point one latitude, +-90 degrees, floating point

P1Lon point one longitude, +-180 degrees, floating point

P1Alt point one altitude in meters, from sea level, -500 to unlimited meters, floating point

P2Lat point two latitude, +-90 degrees, floating point

P2Lon point two longitude, +-180 degrees, floating point

P2Alt point two altitude in meters, from sea level, -500 to unlimited meters, floating point

Time requested time of calculation, DDMMYYYYHHMM or 000000000000,

Each single line contains the input parameters. Latitudes and longitudes are double precision values with N digits of precision.

Output Format:

P1Lat point one latitude, +-90 degrees, floating point

P1Lon point one longitude, +-180 degrees, floating point

P1Alt point one altitude in meters, from sea level, -500 to unlimited meters, floating point

P2Lat point two latitude, +-90 degrees, floating point

P2Lon point two longitude, +-180 degrees, floating point

P2Alt point two altitude in meters, from sea level, -500 to unlimited meters, floating point

Time requested time of calculation, DDMMYYYYHHMM or 000000000000,

ErrorCode return error (refer to error table), integer

For each input line, a record is returned. In summary, the header line repeats the input parameters and adds the ErrorCode.

Latitudes and longitudes are double precision values with N digits of precision. For each header, a single line is output that contains the LEC with the following format:

Ptpec calculated Point-to-point Electron Content, floating point

For multiple position requests, within a single SOAP request, the response will be ordered the same as the request order.

2.4.2. Near Real-Time System Requests

To request results from the most recent model output, use a timestamp that is all zeros (i.e., "000000000000"). This *zero* timestamp feature allows the user to be unconcerned with knowing the exact current time, and to simply request the latest model output results.

It should be noted that near real time results can also be requested by using current timestamps. All timestamps are considered to be in the UTC timezone, and all timestamps are relative to the web services server clock.

If near real time results are requested and data isn't available that is more recent than some threshold of minutes, then the SOAP response will contain the "data not available" error code. The threshold is determined by the frequency of model output and a tolerance factor. Currently, the approximate thresholds are as follows:

Method	Model Output Frequency (in minutes)	Threshold (in minutes)
Tec	60	63
Ustec	15	20

2.4.3. Return Errors (Error Codes)

0 no error, data is being returned

-1 invalid receiver LAT value input or out of range

-2 invalid receiver LON value input or out of range

-3 invalid receiver Altitude value input or out of range

-4 invalid receiver Time value input

-5 data not available, unknown reason

-6 data not available, model data not available at requested time

-7 data not available, satellite data not available at requested time

-8 data interpolation spans incompatible datasets (ie. realtime and assimilated)

-9 can't compute line-of-site (earth blocks Ptpec computation)

-10 point in Lec request is outside of computational domain

These return error codes are not SOAP faults, but application defined errors that signify backend processing errors for each individual (receiver) request. A single returned SOAP fault code indicates that there was a SOAP service error or that the back-end processing failed completely.

2.4.4. Computational Considerations

It should be noted, that for the Tec, Lec, and Ptpec methods, a time-based interpolation is performed for Electron Content calculations between model output data. In other words, the global model creates output on the hour (i.e., HH:00). Thus, for timestamp requests that are not on the hour, the Web Service results will be interpolated. In the future, we will be improving model accuracy, and creating model output every minute.

Interpolation is not used for the Ustec method; instead, the output data used will be the latest data that is output, which is immediately prior to the requested timestamp. Refer to the model output frequencies for each model to better understand how each output dataset is used.

The load that Web Services places on our system is dependent upon several factors. One factor that influences response time to the user is request size (i.e., the number of location/timestamp requests contained in a single SOAP request). While the "back-end" SOAP service can process each location/timestamp request in less than 1 second (even 0.1 seconds), the efficiency of the client process and the speed of the CPU that runs the client also affects the overall response times. Expected response times range is from 1 to 10 seconds for each location/timestamp request and there are size limitations within each client (e.g., 10,000 requests for the Fusion cSoap 2.5 version).

For a given specific location, obtaining a day worth of 1-minute Slant TECs means a request-size of 60*24 = 1440.

The user may need to partition your request size into multiple SOAP requests to obtain an efficient response time or solve client timeout issues. You would need to use a client's –timeout (-t) option to enlarge the client timeout value.

2.5. Real-Time Streaming GPS Data and NTRIP

Our ionospheric specification system has been continuously operational for over 2 years. It was only capable of generating delayed data due to latencies associated with obtaining data from IGS network

receivers. Several dozen IGS stations are generating real time streaming GPS data; however, such data are not readily available publicly. While this situation is likely to change in the future, it appears that recent developments resulted in availability of non-IGS sources of streaming real time GPS data.

In November 2004, the Radio Technical Commission for Maritime Services (RTCM) adopted a new standard for Networked Transfer of RTCM via Internet Protocol (NTRIP). The NTRIP is a non-proprietary HTTP-based open protocol that was initiated by the German Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie, BKG) within the framework of the EUREF-IP Pilot Project. NTRIP clients and servers are publicly available for download (http://igs.ifag.de/index_ntrip.htm). Several NTRIP broadcasters that re-transmit streams of GPS data from a number of regional networks free of charge are already operating: http://igs.ifag.de/ntrip_caster.htm , http://igs.ifag.de/root_ftp/misc/ntrip/maps/All-World.gif

The rationale for developing NTRIP is given by the following RTCM press release:

"Global Navigation Satellite Systems (GNSS) provide geographical positioning information from a constellation of satellites in orbit to receivers at sea, on the ground, and in the air. The best known of these systems is the U.S. Global Positioning System (GPS), but the Russian GLONASS system provides a similar service, as will the European Galileo system. Together they are known as Global Navigation Satellite Systems, and they can provide position accuracies in the 10-meter to 15-meter range. Although the satellites have the potential to provide more accurate positions, atmospheric and other effects degrade the quality of the satellite signals. As impressive as GNSS systems are, they do not directly provide accuracies that are good enough to rely on for ships entering harbors, or docking, for example. The satellite signals can be corrected by using reference stations at precisely known locations which broadcast corrections to GNSS receivers nearby. This technique is known as Differential GNSS (DGNSS) service, and it has enabled precise navigation not only by ships, but also aircraft, and ground vehicles. Centimeter level precision can now be obtained, allowing tractors to cross agricultural fields in precisely the same track every time, improving crop yields and enabling snow plows to operate quickly over roads buried beneath an otherwise trackless snow field. New applications continue to be developed. Typically, differential corrections have been broadcast over radio data links from single reference stations located in precisely known locations, to mobile receivers (rovers) located on the equipment whose position needs to be known. As the uses of DGNSS services have grown, governments and commercial service providers have established networks of reference stations. One way to further increase accuracy is to use correction data from multiple reference stations, such as these networks provide. For all these applications, replacing the radio data link with data streaming over the Internet to stationary or mobile users using the Ntrip protocol, can be advantageous.

The Ntrip project was initiated by the German Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie, BKG). Although there are uses for stationary DGNSS receivers that could access the Internet via landline, the growing availability of Internet service through the mobile telephone network was a persuasive reason to develop and formalize a publicly available Internet protocol for streaming DGNSS data. Ntrip is designed to distribute differential correction data or other kinds of GNSS streaming data to stationary or mobile users over the Internet, allowing simultaneous PC, Laptop, PDA, or receiver connections to a broadcasting host. Ntrip supports wireless Internet access through Mobile IP Networks like GSM, GPRS, EDGE, or UMTS. Ntrip is meant to be an open non-proprietary protocol. Major characteristics of Ntrip's dissemination technique are the following:

- It is based on the popular HTTP standard, and is comparatively easy to implement when limited client and server platform resources are available.

• Its application is not limited to one particular plain or coded stream content; it has the ability to distribute any kind of GNSS data.

• It has the potential to support mass usage; it can disseminate hundreds of streams simultaneously for up to one thousand users when applying modified Internet Radio broadcasting software.

• Regarding security needs, stream providers and users are not necessarily in direct contact, and streams are usually not blocked by firewalls or proxy servers protecting Local Area Networks.

• It enables streaming over any mobile IP network using TCP/IP."

Currently, there are approximately 30 operational broadcasters and this number is growing. The broadcasters redistribute raw GPS data streams that they receive as NTRIP streams. BKG provides access to a number of open source software clients that can be used to obtain these data with very small latencies. A number of real time steams worldwide have been made available to us by BGK.

We believe that in the very near future there is a significant potential for the emergence of a new class of global real time sources of GPS data from regional networks of reference stations connected to the Internet. In our view, this development is driven by a number of factors:

- The so called Real Time Kinematics (RTK) systems rely on a regional or local network of continuously operated GPS reference stations to enable differential GPS positioning accuracy of several centimeters in near real time. The networks are needed to account for and eliminate local systematic errors due to ionospheric and tropospheric effects and help to resolve integer ambiguities for carrier phase positioning. These systems enable significant cost savings in areas such as precision agriculture, assisted steering, surveying, maintenance, snow plowing, and others. On the other hand, driven by demand, equipment prices appeared to have reached a level when acquiring, installing, and operating such networks is within the capabilities of a number of organizations worldwide.

- Big manufacturers such as Trimble and Leica like to use this opportunity to further promote RTK networks and push their equipment and services.

- Traditionally, corrections from reference stations in RTK networks are delivered to a roving receiver(s) via FM or AM radio link. This requires additional infrastructure and often a direct line of sight between the rover and a reference station. In addition, it normally necessitates a 1-way communication and therefore a need to (often wastefully) broadcast corrections for the whole grid covered by the network. Proliferation of GSM cellular networks and other cellular networks that allow Internet access via cell phones resulted in a cheap, convenient, and relatively reliable way for 2-way communication between multiple rovers and RTK network stations..

- It would appear that even with new civilian frequencies and Galileo functioning, instantaneous reliable integer ambiguity resolution in RTK scenario will only be possible for very short network baselines. Thus, adding new frequencies and satellites will not (or so it seems at the moment) eliminate the need for such networks.

While researching availability of real time streaming data for assimilation to our system, we discovered that Internet-connected GPS networks exist in a number of regions: Europe (vast majority), South America, Australia, South Africa, South Korea, China, Hong Kong, Japan, New Zealand and others. Other possible sources of real time streaming data are the 2 providers of world-wide DGPS

service; namely, NAVCOM Technology (http://www.navcomtech.com/starfire.cfm) and Omnistar (http://www.omnistar.com/).

Data from some of these networks might be available for a fee. A number of the abovementioned NTRIP broadcasters make data available free of charge with no guarantees of availability or reliability.

One should keep in mind that not all streams available from NTRIP broadcasters contain information relevant for ionospheric electron content determination. Some of these streams can only contain local position correction and not the raw code and phase measurements.

In the process of executing our current Phase II SBIR contract, we signed agreements with GPS streaming data providers from Korea, Australia, and Europe. We can now receive near real-time (with latencies of ~ 10 seconds) raw 1Hz GPS data from a number of locations worldwide.

We developed capabilities for converting these streams to RINEX files and deriving slant total electron content from these data. We are in the process of incorporating these streaming data into our assimilation system.

2.6. Scintillation Growth Rate Diagnostics

Equatorial scintillations or plasma bubbles are small scale perturbations of the ambient electron densities that, under certain conditions, can rapidly grow and result in partial or complete fading of the GPS signal. While some dual-frequency receivers have capabilities to mitigate their effect (sometimes in real time and sometimes via post processing), it is likely that strong enough scintillations will disrupt operations of any receiver.

Availability of new GPS and GALILEO satellites will improve chances that for a given receiver there will be enough satellites whose signals are not affected by scintillations. Yet, especially in the equatorial regions, there will likely exist regions where all useful signals are blocked for a period of time.

Ionospheric scintillations can also interfere with other ground-to-satellite and ground-to-ground communication channels. Therefore, the capability to forecast scintillations and warn the end user can be rather valuable.

One can separate approaches to modeling and forecasting scintillations into two major classes: explicitly resolving growth and evolution of these instabilities with very high-resolution fluid dynamics codes; and diagnosing favorable conditions for instability growth and thus forecasting favorable conditions for occurrence of scintillations at a given time and location.

We argue that the first approach is likely not to be practically feasible on a global scale for a number of years, if ever, due to the necessity to resolve very small spatial scales. The "probabilistic" approach (e.g., Sultan 1996 and Secan et al. 1995) relies on ambient properties of the ionospheric plasma such as ion and electron densities, and density gradients, ion and density collision frequencies, recombination rates, inclination and declination of the magnetic filed lines, ion velocities and temperatures, etc., to compute the linear growth rate of plasma instabilities as a function of geographic location and time. The greater the growth rate, the higher the chance of severe scintillations. Since all the necessary parameters are routinely computed in this model, it is relatively straightforward to diagnose the scintillations' growth rates. In forecast mode these rates can be predicted.

We implemented a module for calculating the scintillation growth rates. An example of the calculated average growth rates is shown in Figure 6. Regions colored in dark red have favorable conditions for occurrence of scintillations. Maps similar to the one shown in Figure 6 are routinely generated by the system and displayed at http://fusionnumerics.com/ionosphere .

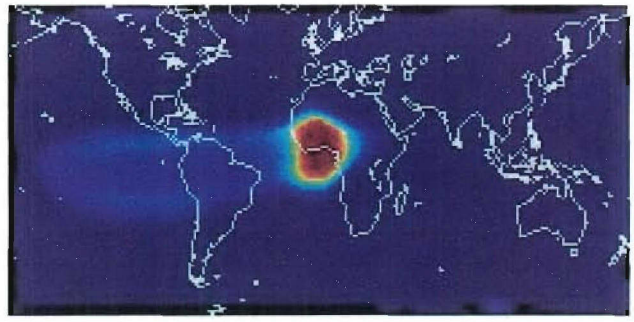


Figure 6. A map showing diagnosed scintillation growth rates.

2.7. Longer-Term Forecasts

Longer-term (12 hours to several days) forecasts of ionospheric conditions require forecasting solar flux and solar wind parameters (speed and magnetic field) near the Earth as a result of solar activity. A rapid increase in the southward component of the solar wind magnetic field normally leads to geomagnetic storms. It is common knowledge that large geomagnetic storms (e.g., October-November 2003) result in rapid and strong increase in the spatial extent and strength of scintillations.

As a part of this project, we have researched implementation of a forecasting module that analyzes time sequences of solar images obtained by SOHO satellite using an artificial intelligence technique known as Support Vector Machines (SVM). The system is a classifier trained on a multi-year data set of past events and solar images. While it does not aim to forecast the exact state of solar wind, it achieves reasonable success classifying a particular sequence of solar images, either as likely to be leading to a geomagnetic storm or resulting in undisturbed conditions.

SOHO EIT images are automatically downloaded through the EIT catalog webpage every half-hour, when available. Every six hours, these raw images are calibrated using eit_prep and solarsoft software packages obtained from NASA. Calibration involves flatfielding (setting a background intensity level), removing grid artifacts that are generated when the picture is taken and missing block removal.

All data is archived on a daily basis, and we have a full set of all 96,000 (or so) EIT images in raw and calibrated form. At the same time the calibration is done, an attempt is made to find potential coronal mass ejections (CME) in the EIT images using a support vector machine. For this prediction, there is a two-stage process.

First, in order for SVM to be effective, it must be trained. The training process is undertaken each time a new event is found and at the start for the whole set of preceding images.

This starts with an analysis of ACE data to identify geo-effective events. Once a set of events has been identified, a corresponding input vector is made that is a time sequence of differenced images, centered at the time when an event is determined to have occurred at the sun. The actual event times at the sun are determined by travel-time analysis using solar wind speed conditions which are accurate up to within roughly 11 hours of the actual time. For further accuracy, the pair of images that changed the

most (via the norm of the difference vector) within a half-day of the travel-time calculation is taken to be the true center of the event. This is based on an assumption that events are accompanied by flares, if not exactly at CME liftoff, then nearby in time.

About 80 major events have been identified for 2000-2003 via the automated method. For reasons of compatibility problems between 512 and 1024 images, a model is made for each type of image separately, which means we have coverage of roughly 30 and 50 events for 512 and 1024, respectively.

In order to train a model for prediction, SVM requires a number of non-event vectors as well, so an equal-sized set of non-event vectors is made using quiescent ACE data. SVM is trained on this set with standard cross-validation methods and the resulting model is used for prediction.

After the model has been trained, new images are analyzed with SVM for the likelihood that an event occurred. This is done by using SVM to classify a set of three-hour-long vectors that represents roughly one day of data. If an event is identified, it's time is recorded for further analysis.

We have also explored a complimentary approach to image classification. The idea is to identify hot spots on a solar image, and characterize them by computing a variety of properties to form an 8-dimensional vector. We identify hot spots by extracting all pixels about a given intensity threshold, however, it is a parameter. Next, we compute the connected components of these regions. For each region, we compute:

- 'Area' - The actual number of pixels in the region.
- 'Centroid' – The center of mass of the region in (x,y) pixel coordinates.
- 'MajorAxisLength' - The length (in pixels) of the major axis of the ellipse that has the same second-moments as the region.
- 'MinorAxisLength' –The length (in pixels) of the minor axis of the ellipse that has the same second-moments as the region.
- 'Eccentricity' - Scalar; the eccentricity of the ellipse that has the same second-moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1. (0 and 1 are degenerate cases; an ellipse whose eccentricity is 0 is actually a circle, while an ellipse whose eccentricity is 1 is a line segment.)
- 'Orientation' - Scalar; the angle (in degrees) between the x-axis and the major axis of the ellipse that has the same second-moments as the region.
- 'Solidity' - Scalar; the proportion of the pixels in the convex hull that are also in the region.
- 'Extent' - Scalar; the proportion of the pixels in the bounding box that are also in the region.
- Average Intensity – The average of the pixel values in the region.

Currently, each hotspot is classified as an event if the hotspot occurs in an event image and a non-event otherwise. Clearly this misclassifies many rather innocuous hotspots as events.

For several different types of events, we were able to create an SVM model that could identify the event vectors with at least 75-percent accuracy and often above 90-percent, given a set of equal numbers of event vectors and random non-event vectors.

The greatest hindrance to the task appears to be a lack of recorded events and the incompatibility of EIT images of 512- and 1024-pixel resolutions. Until we can reliably normalize the intensity values of these two image types, we cannot train the SVM on more than one type at a time. This means there are 45 events between 2000 and 2004 for which we have sufficient data coverage to train a model for a given image type. Moreover, when attempting to train SVM on sets weighted to include more non-events than events, as would be the case in real time prediction, the ability to find events fell off very quickly.

We established event lists based on three ACE parameters: Bz*|Vx| (the z component of the magnetic field passing through L1 times the speed of the field in the direction of Earth), total magnetic field strength and bulk wind speed. For Bz*|Vx|, an event was considered significant negative values lasting more than two hours, which implies a substantial portion of the Earth's magnetic field could be eroded. Field strength and bulk wind speed parameters were used to track shock waves in the interplanetary magnetic field that often precede CMEs. Non-events were taken as quiescent periods lasting at least eight hours.

In all three cases, the event lists included clearly identifiable events, but they also identified events as periods of disturbance in a given parameter that did not relate to a recognized CME. For this reason, these lists were not reliable for training SVM models. Nonetheless, cross-validation for lists' corresponding vectors, encoded by the method described below, resulted in event recognition accuracy of more than 75 percent, with a 15-percent false-positive rate, for sets of an equal number of events and non-events.

We attempted pseudo-real time event prediction, whereby an SVM model was generated and used to search for known events in past data. Time-consecutive sets of images were fed to SVM for classification, as if it were tracking input in real time. The results were disappointing, however. Either the model predicted an event for almost every set of vectors it analyzed, or no events were found for any vectors. In addition, we deployed a real time prediction process, but it did not do any better than the pseudo attempt.

Although there is no guarantee that the method we are using to encode the images for SVM training is effective, there is a clear problem of a lack of events for which we have data coverage. Using the full set of over 96,000 E.I.T. images, we could encode vectors for roughly 100 of the 150 events. Roughly half of these are in 512-pixel resolution images and half are in 1024 images.

According to a personal communication from Dr. Joe Gurman, chief scientist for the SOHO project, 512- and 1024-pixel resolution images cannot be normalized, thus, their values are essentially incompatible. This means SVM can only be trained for one type of image at a time. Therefore, for any given model, there are at most 50 encoded events in the time period of 2000 to 2004. This is clearly a major obstacle to developing a reliable classification scheme.

3. RESULTS AND VALIDATION OF THE IONOSPHERIC MODEL

To analyze the assimilation system performance we conducted a number of standard statistical tests. We calculated the post-fit residuals for all stations used in the assimilation process. These post-fit

residuals are root mean-squared differences between the slant TECs that were derived from data with receiver biases removed, and the slant TECs after the Kalman filter analysis "blended" model results and data. Note that data for all stations are assimilated simultaneously and that the employed Kalman filter assumes that data from each station influences electron densities within a certain distance from its location, however slightly. This distance can be as large as thousands of kilometers, depending on the station location. Therefore, small errors in the neighboring station data originated from non-optimal receiver bias estimation, or errors created by the phase-leveling algorithm, can negatively affect the results since the input data are not self-consistent.

Assuming that the Kalman filter is implemented correctly, the magnitude of post-fit residuals is indicative of (1) the theoretical best system performance in its current configuration, and (2) the accuracy of the ionospheric delays for short baselines (comparable to the spacing between model grid-points, ~100-300 km).

The residuals' statistics are presented in Figure 7. The maroon line shows post-fit residuals and the blue line shows pre-fit residuals averaged over all station-satellite pairs. The difference between the two curves indicates an accumulated error (~0.5 TEC units) due to the 10-minute model forecasting step. A new analysis is performed every 10 minutes and the plot corresponds to approximately 3 days.

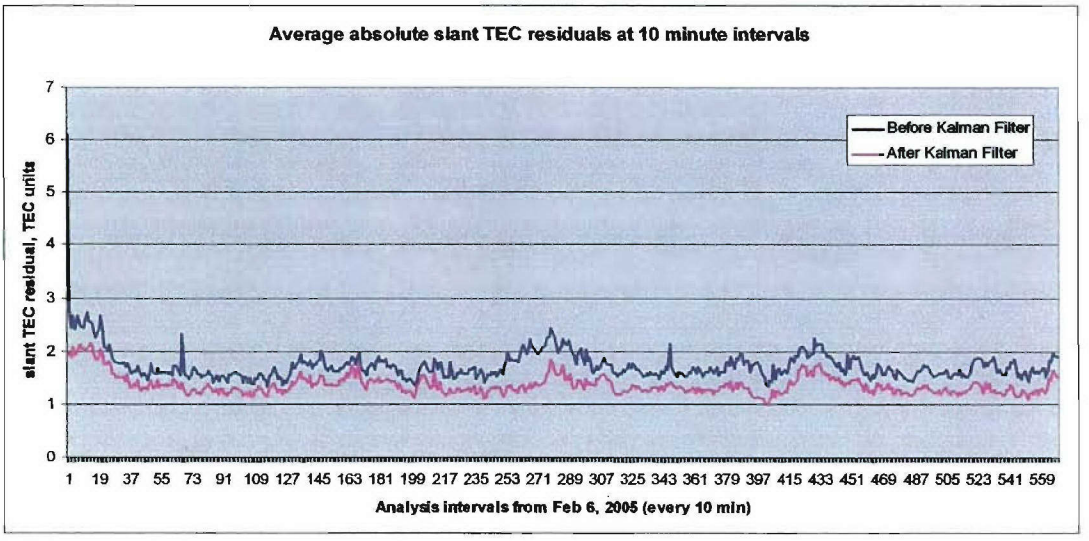


Figure 7. An example of globally averaged post- and pre-fit residuals at 10-min intervals.

Another useful test of the assimilation system is the so called χ^2 test, described, for instance, in Talagrand (2003) or Menard et al. (2000). Put simply, χ^2 is computed as an average value of the ratio of the directly computed observation-minus-forecast root mean-square error and forecast error diagnosed by the Kalman filter. As such, if tuneable parameters of the Kalman filter implementation are chosen approximately correctly, the average value of χ^2 should be 1. Figure 8 shows a time series of normalized χ^2 produced within an assimilation experiment with filter parameters tuned using the procedure described in Khattatov et al. (2000). After the adjustment period, the average value of χ^2 becomes approximately one and it does not exhibit a time trend.

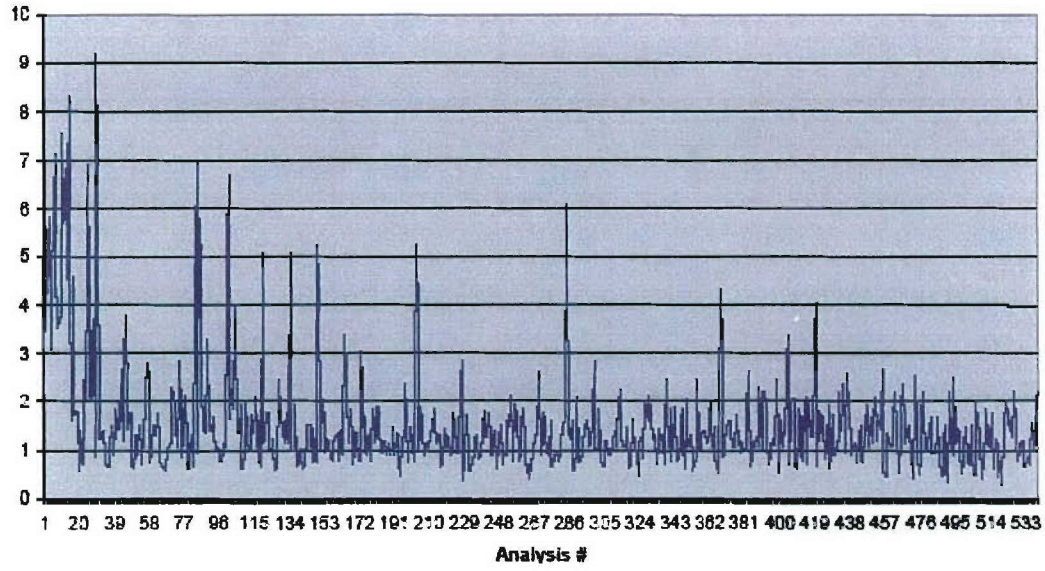


Figure 8. $\chi 2$ test results.

A randomly selected subset of 5 GPS stations is withheld from the assimilation process for 72 hours. These stations are designated as control or "ground truth." During these 72 hours we compute slant TEC values from these control stations as well as slant TECs for the locations of these stations predicted by the model. At the end of the 72-hour period, a new subset of control stations is chosen at random from the remaining stations that have been used for assimilation and the current control subset is returned to the assimilation pool. This methodology allows us to systematically estimate system errors for all stations. Note that the 72-hour time window is sufficiently longer than the typical ionospheric response time of just several hours. Therefore no "memory" of the control station measurements remains in the system. Similarly to common practices at operational meteorological centers, we intend to continuously compile these long-term statistics.

Figure 9 shows an example of time series of slant TEC generated in the system and measured by a control GPS station located in North America. Due to the relatively high density of GPS network in that region and generally fairly small ionospheric gradients, the system's accuracy is quite good.

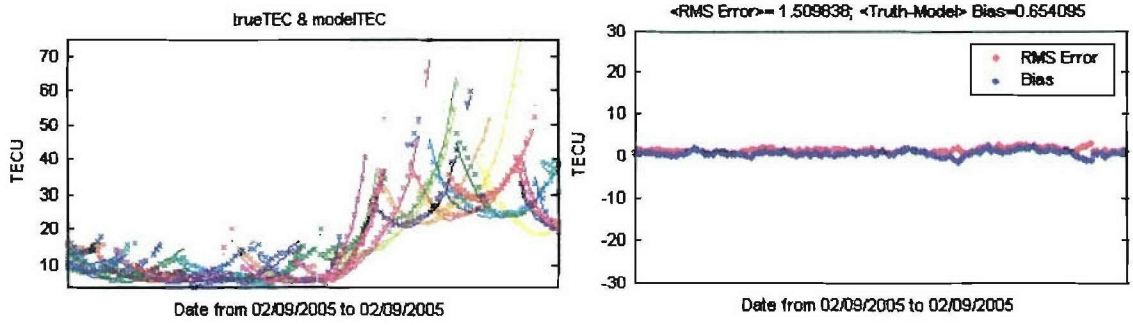


Figure 9. Left: An example of time evolution of slant TEC in TEC units from the assimilation system (solid lines) and control GPS reference station measurements (dots) to several GPS satellites in view. Different colors correspond to different satellites. Right: Absolute slant TEC errors and biases in TEC units as a function of time of day.

Figure 10 shows the cumulative performance statistics from May 1, 2005, through July 10, 2005, (the time of this writing). Distribution of this ~2.5-month-long period each used IGS GPS station has been

selected at least once for validation. Therefore these results are representative of the mean global performance over this time period. As shown in Figure 10, the global mean system bias is approximately zero, the global RMS error is about 5 TEC units. If only stations located Northward of 20N are considered, the RMS error decreases to 3 TEC units due to the higher density of reference stations in the Northern Hemisphere.

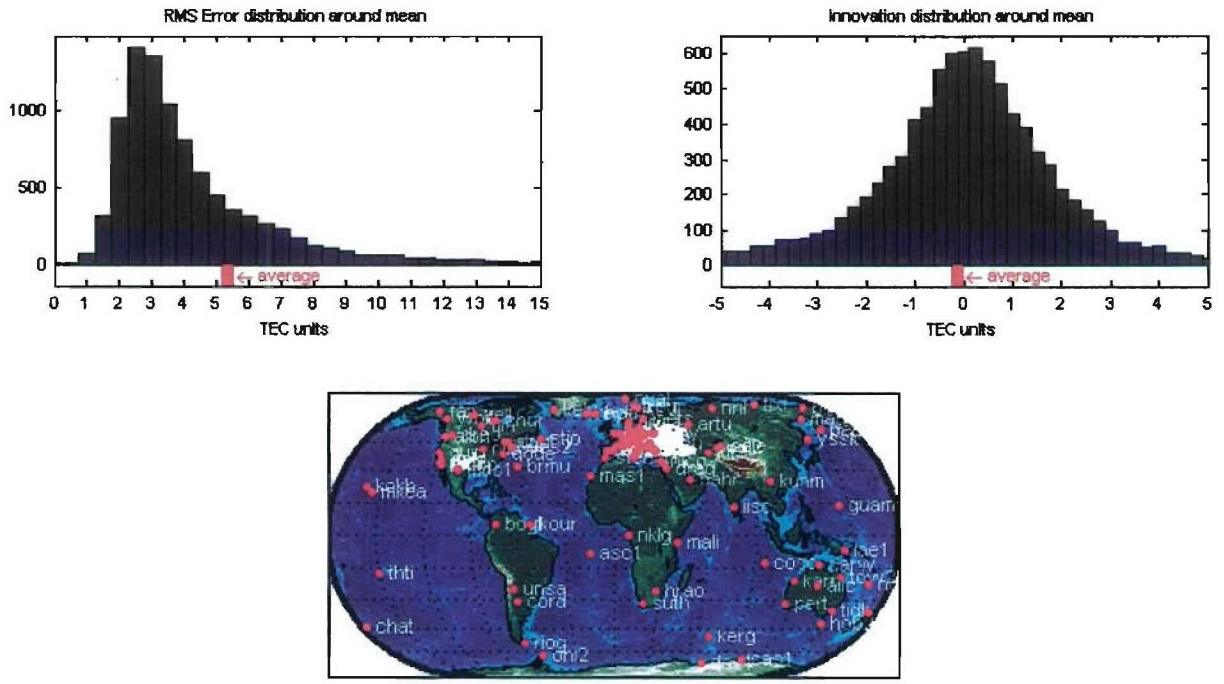


Figure 10. Cumulative global distribution of RMS errors (top left), system biases (top right), and validation stations (bottom) for May 1 – June 10, 2005, time period.

3.1. Comparisons with Independent Data

The ultimate test of any assimilation system is systematic comparisons with independent data. Assimilation results are being currently downloaded to the National Geophysical Data Center and compared operationally with dynasonde data from the Bear Lake Observatory. The dynasonde, together with its automated data analysis system, obtains local electron density profiles, velocities, and derived parameters. It is likely to play an expanding role in Space-Weather assimilative modelling. Presently, we use only the profiles and derived TEC estimates, and we use them not for assimilation but for independent comparison and validation only. This strategy acknowledges that ionosonde data, in common with most geophysical observations, have their own validation and error-estimation problems. An evolutionary process is planned to coincide with deployment of new state-of-the-art dynasondes and their advanced "Dynasonde-21" data analysis and data distribution system now in a late stage of development. Most of this data system is already functional and can be accessed for the legacy-dynasonde installations of Bear Lake, EISCAT (Norway), and Lycksele (Sweden) at: http://www.ngdc.noaa.gov/stp/IONO/Dynasonde/.

Dynasonde electron density profiles are obtained by a new three-dimensional inversion procedure, NeXtYZ (Zabotin, et al. 2005). The bottomside TEC is obtained simply by integration of the profile up to the F-region peak. At and above the peak, the profile is represented by a "Chapman function", which is the appropriate form according to standard theory for a plasma layer in diffusion/loss equilibrium

(Wright 1960); the scale height H, peak height hmax, and peak density, Nmax, which define this function are determined in the inversion process. The scale height, H, is assumed constant with height on the topside, but this is only an approximation. At a later stage of our collaboration, an estimate of dH/dh from the assimilation system might be used to adjust the dynasonde TEC for an improved comparison. Since mid-2004, results from the assimilation system have been systematically downloaded to the NGDC every 15 minutes. There, they are archived and prepared in graphical form for comparison with the dynasonde data as may be seen at (http://www.ngdc.noaa.gov/stp/IONO/Dynasonde/IonoNumerics_a.htm). These comparisons allowed us to quickly assess effects of various improvements to the system and were crucial in locating and eliminating a number of software bugs. There are two types of comparisons: (1) system-generated electron density profiles above Bear Lake are converted to plasma frequency and displayed for comparison with the dynasonde profiles; (2) assimilation total vertical electron content is displayed next to the dynasonde-estimated TEC as a function of time.

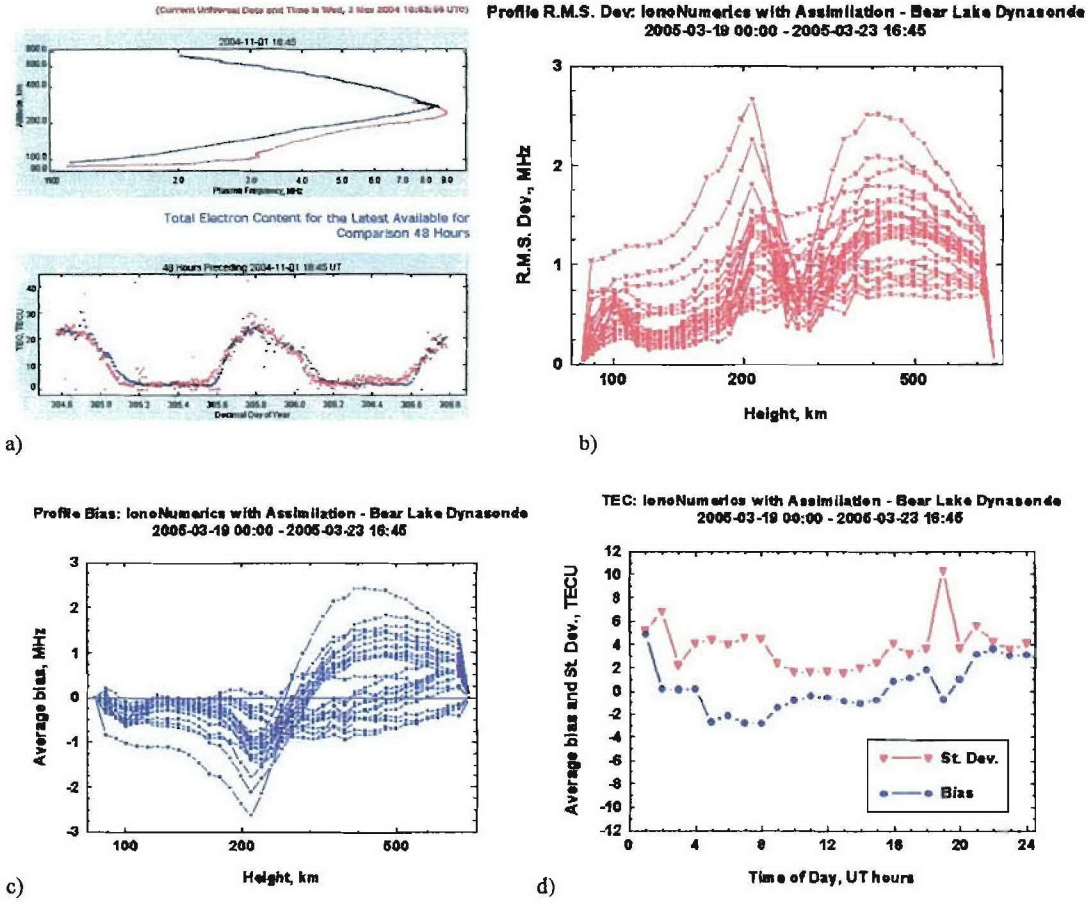


Figure 11. (a) Example comparisons of plasma frequency profile and total vertical electron content from Bear Lake dynasonde (red) and Fusion Numerics assimilation results (blue); (b) Root-mean-square differences between assimilation results and the dynasonde data expressed in MHz as a function of altitude; (c) Systematic differences (biases) between assimilation results and the dynasonde data expressed in MHz as a function of altitude; (d) Root-mean-square differences (red) and biases (blue) between assimilation results and dynasonde data for total vertical electron content expressed in TEC units as a function of time of day.

Figure 11(a) is an example of these comparisons with Fusion Numerics' results in blue and dynasonde data in red. General features of the electron content are similar in both time series. Moreover, the difference between the dynasonde-estimated TEC and the assimilation model is usually

only about several TEC units, which is rather encouraging. Figures 11(b) and 11(c) demonstrate root mean square differences and systematic differences (biases) between dynasonde and the assimilation system for the period between March 19 and March 23. Currently such statistics are generated as a 7-day running average and are available for monitoring and documenting assimilation system performance. Finally, Figure 11(d) shows RMS and biases for the total electron content from both sources of data as a function of time. As one can see, both parameters are near 2-4 TEC units in agreement with statistics obtained from our control IGS study.

Due to the complexity of the equatorial ionospheric processes and generally much higher plasma densities, one would expect comparisons with data near the magnetic equator to be more challenging. Figure 12 shows electron density profiles obtained at the Jicamarca radio observatory located at the magnetic equator and corresponding profiles from the assimilation system. The Jicamarca data has been processed with the technique described in Feng et al. (2004). While Jicamarca measurements are not used in the system, the agreement between data and the assimilated profiles is rather encouraging.

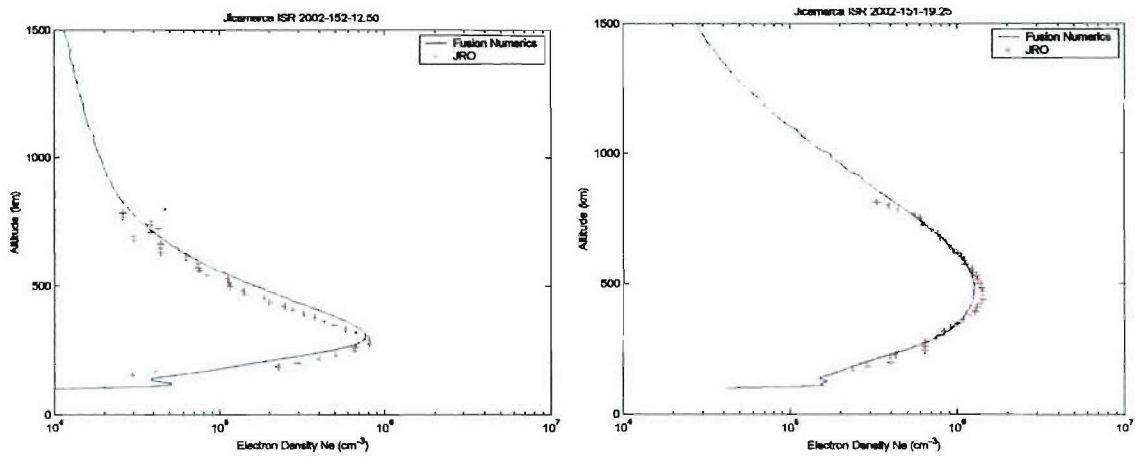


Figure 12. An example of altitude profiles of electron density from Jicamarca radion observatory (crosses) and the assimilation model. The Jicamarca data were provided by J. Chou and D. Hysel. Jicamarca data are not used in the assimilation.

We have made arrangements with a web-site operator that provides real time propagation information for amateur radio operators around the world to display "live" real time 6-m VHF communication links superimposed by TEC maps generated by us. An example is shown in Figure 13. Black lines show the established radio links that are likely a result of the so-called hordial jump when the emitted signal reflects first from the closest "tail" of the 2-prong region of enhanced electron density in the east-west direction. It then bounces off of the second tail and gets reflected downward where it reaches the receiver. These results are available on-line in real time at http://maps.dxers.info/tec/6m_tec.php and provide an additional confirmation of the realism of the generated TECs.

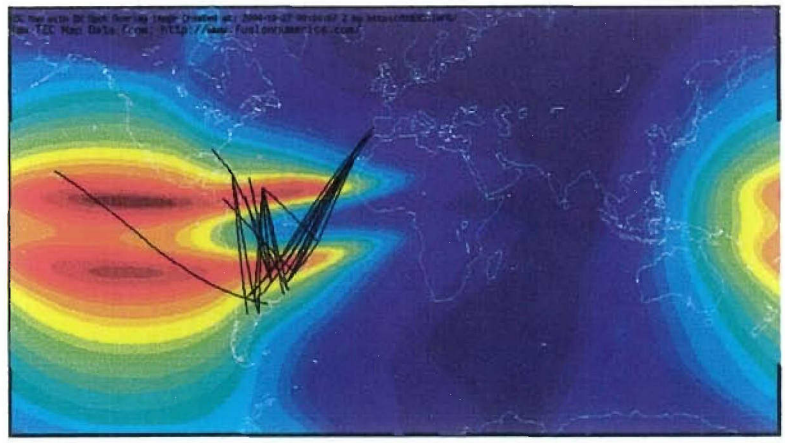


Figure 13. An example of established 6-m radio communication links superimposed onto a map of total electron content generated in the system.

4. CONCLUSIONS

This report presents results of an AFRL-sponsored project to develop advanced modeling and data assimilation capabilities for the ionosphere and upper atmosphere. In the course of this project, Fusion Numerics developed a new global three-dimensional numerical model of the ionosphere, ionospheric data assimilation software and methodology, and an infrastructure for distributing results of the assimilation system.

We believe that the designed system is unique in its numerical implementations of ionospheric physics, data assimilation, and a modern disciplined approach to software engineering. The developed ionospheric modelling and assimilation system solves momentum, energy, and mass conservation equations, and implements assimilation in model internal magnetic coordinates. An operational prototype of the system has been continuously working since August 2003. Preliminary results indicate an absolute globally averaged assimilation error of 2-5 TEC units and relative error of approximately 5-10%. These results indicate that the pursued approach is viable and has practical merit.

Recently, we have also developed a module to evaluate equatorial plasma-instability growth rates using the formalism of Sultan (1996); its results are available online. Other applications of the full system include its use for radio propagation ray-tracing (Angling and Khattatov 2005). Comparisons of the ray-tracing results obtained with different ionospheric models helped us to identify and eliminate a number of early system deficiencies. Forecasting capabilities are currently being developed via implementation of ensemble Kalman filtering. In the course of the new development effort, additional features will be added to the system and its performance is expected to improve.

Accuracy and precision depend strongly on geographic location and are influenced by the reference station density. Note that these results are obtained with a fairly crude phase-leveling mechanism that skips 20 minutes of data every time a cycle slip occurs for a given receiver-satellite pair. Fairly straightforward enhancements to the system should result in better accuracy. Receiver bias estimation will also become more accurate for longer initialization periods.

We expect the most significant jump in performance from assimilating reference network input data more often; for example, every minute instead of every 10 minutes. The necessary modifications in the processing methodology are trivial and the implementation depends simply on faster computing capabilities as the assimilation step takes significant memory and CPU resources.

5. REFERENCES

Angling, M., and Khattatov, B., Comparative study of two ionospheric imaging techniques. This symposium, 2005.

Bailey and Balan, A low-latitude ionosphere-plasmosphere model, in STEP: Handbook of Ionospheric Models, STEP Report, editor R.W. Schunk , 1996.

Blewitt, G. (1990): An Automatic editing algorithm for GPS data. , *Geophys, Res Letters*, vol. 17, No. 3, pp 199-20

Fejer, B. G., and L. Scherliess (1995): Time dependent response of equatorial ionospheric electric fields to magnetospheric disturbances, *Geophys. Res. Lett.*, **22**, 851-854.

Feng, Z.; Kudeki, E.; Woodman, R.; Chau, J.; Milla, M., F-region plasma density estimation at Jicamarca using the complex cross-correlation of orthogonal polarized backscatter fields, *Radio Sci.*, **39**, 2004

Fuller-Rowell T.J., D. Rees, S. Quegan, R.J. Moffett, M.V. Codrescu, and G.H. Millward (1996): A coupled thermosphere ionosphere model (CTIM). Handbook of Ionospheric Models, STEP Report, editor R.W. Schunk.

Hedin, A. E., E.L. Fleming, A.H. Manson, F.J. Scmidlin, S.K. Avery, R.R. Clark, S.J. Franke, G.J. Fraser, T. Tsunda, F. Vial and R.A. Vincent, Emperical Wind Model for the Upper, Middle, and Lower Atmosphere, J. Atmos. Terr. Phys., 58, 1421-1447, 1996

Hedin, A.E., Extension of the MSIS Thermosphere Model into the Middle and Lower Atmosphere, J. Geophys. Res., 96, 1159, 1991

Huba et al. (2000): SAMI2 is another model of the ionosphere: a new low-latitude ionosphere model, *J. Geophys. Res.*, **23**,035.

Khattatov, B.V., J.-F. Lamarque, L. V. Lyjak, R. Menard, P. F. Levelt, X. X. Tie, J. C. Gille, G. P. Brasseur, (2000): Assimilation of satellite observations of long-lived chemical species in global chemistry-transport models, *J. Geophys. Res.*, **105** , 29135.

Khattatov et al., Ionospheric Corrections from a Prototype Operational Assimilation and Forecast System, *Proceedings of IEEE Position, Location, and Navigation Symposium, (PLANS)*, Monterrey, CA, April 26-29, 2004a.

Khattatov et al., Advanced Modeling of the Ionosphere and Upper Atmosphere, *Final Scientific and Technical Report AFRL-VS-HA-TR-2004-1129*, Air Force Research Laboratory, Hanscom AFB, MA, 2004b.

Khattatov et al., An Ionospheric Forecasting System, *Proceedings of Institute of Navigation Annual Technical Meeting*, Long Beach , California, September 2004c.

Lorenc, A.C., 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177-1194.

Mannucci, A. J., B. D. Wilson, D. N. Yuan, C. H. Ho, U. J. Lindqwister and T. F. Runge, 1998, A global mapping technique for GPS-derived ionospheric total electron content measurements, in Radio Science, Vol. 33, pp. 565-582, 1998.

Millward et al. (1996): A coupled thermosphere-ionosphere-plasmosphere model (CTIP), in STEP: Handbook of Ionospheric Models, STEP Report, editor R.W. Schunk .

Ménard, R. and Chang, L.-P., 2000. Stratospheric assimilation of chemical tracer observations using a Kalman filter, part II: Chi-squared validated results and analysis of variance and correlation dynamics. *Mon. Weather Rev.*, **128**, 2672-2686.

Parkinson, B. W., Spilker, J.J. Jr., Global Positioning System: Theory and Applications, vols. 1 and 2, American Institute of Aeronautics, 370 L'Enfant Promenade, SW, Washington, DC, 1996.

Radio Technical Commission for Maritime Services, RTCM Recommended Standards for Networked Transport of RTCM via Internet Protocol (Ntrip), Version 1.0, *RTCM Paper 200-2004/SC104-STD*, Arlington, VA, 2004.

Schunk and Nagy (2000): The Ionospheres, Cambridge University Press, 2000.

Secan, J., R. Bussey, E. Fremouw, and S. Basu , An improved model of equatorial scintillation, *Radio Sci., 30(3)*, 607–618, 1995.

Sultan, PJ, Linear theory and modeling of the rayleigh-taylor instability leading to the occurrence of equatorial spread F, *J. Geophys. Res., 101*, 26875, 1996

Talagrand, O., 2003. *A posteriori* validation of assimilation algorithms. In *Data Assimilation for the Earth System*, NATO ASI Series, eds. R. Swinbank, V. Shutyaev, and W.A. Lahoz. Kluwer.

Weimer, D. R. (1996): A flexible, IMF dependent model of high-latitude electric potentials having "space weather" applications, *Geophys. Res. Lett.*, **23**, 2549.

Wright, J.W. A model of the F-Region above hmaxF2, J. Geophys. Res. 65, 185-191, 1960

Zabotin, N. A., J. W. Wright and G.A. Zhbankov, New approaches in the problem of electron density inversion from dynasonde measurements using three-dimensional echolocation data, proceedings of the 11th Ionospheric Effects Symposium, Alexandria, VA, 2005